ClearSkies: A Preliminary Study of Gaze-Mapped Scene Segmentation in Training Aircraft Cockpits

Sebastian Oes
sebastian.oes@unisg.ch
University of St. Gallen
St.Gallen, Switzerland

Kenan Bektaş*
kenan.bektas@unisg.ch
University of St. Gallen
St.Gallen, Switzerland

Jannis Strecker-Bischoff
jannisrene.strecker@unisg.ch
University of St. Gallen
St.Gallen, Switzerland

Simon Mayer
simon.mayer@unisg.ch
University of St. Gallen
St.Gallen, Switzerland

Figure 1: Flight instrument segmentation using Detectron2 (left), *SAM 2* (middle), and *SAM3* (right) in an Aquila AT01 aircraft.

ABSTRACT

In pilot training, deviation from standard procedures is a significant concern. To provide student pilots with objective feedback in post-flight debriefing, we captured pilots' view and gaze with the Pupil Core eye-tracker. Then we conducted a *preliminary* evaluation to test the feasibility of existing scene segmentation models for gaze-mapping[1]. We used an OpenCV baseline model for coarse inside vs. outside-analysis, a fine-tuned Detectron2 model for specific instrument segmentation, and *Segment Anything Models* (*SAM 2* and *SAM 3*) for human-in-the-loop analysis. The baseline was fast but fragile, failing in common flight scenarios; the Detectron2 model was powerful but inflexible and unsuitable for general use; and *SAM 3* was promising, offering generalizability for post-flight analysis despite noisy digital displays. A qualitative preliminary evaluation of SAM with *Visual Flight Rules* shows that it can be beneficial in eye movement analysis. We identified poor data quality in bright cockpit environments and ergonomics as main limitations.

KEYWORDS

scene segmentation, computer vision, eye tracking, pilot training

*Corresponding Author: kenan.bektas@unisg.ch
[1]https://github.com/Interactions-HSG/ClearSkies

ETRA '26, June 01–04, 2026, Marrakesh, Morocco

ACM ISBN 979-8-4007-2519-7/2026/06.
https://doi.org/10.1145/3797246.3805852

ACM Reference Format:
Sebastian Oes, Kenan Bektaş, Jannis Strecker-Bischoff, and Simon Mayer. 2026. ClearSkies: A Preliminary Study of Gaze-Mapped Scene Segmentation in Training Aircraft Cockpits. In *2026 Symposium on Eye Tracking Research and Applications (ETRA '26), June 01–04, 2026, Marrakesh, Morocco.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3797246.3805852

1 INTRODUCTION

The use of eye tracking technology in aviation has seen growing interest since initial studies dating back to the 1950s [Fitts et al. 1950]. In pilot training, eye tracking may provide insight into the visual behavior of instructor and student pilots and allows them to provide objective feedback about their behavior and situation awareness [Niehorster et al. 2020]. In such *diagnostic* applications of eye tracking [Duchowski 2018], pilot's gaze data need to be mapped to specific areas of interest (AOIs), which can be determined manually either before or after data collection. To minimize processing and analysis time in eye-tracking studies, it is necessary to transition to automated solutions [Ferrari et al. 2018]. For example, computer vision techniques can be used to automatically segment AOIs in a given video feed, which then helps to distinguish whether the pilot was looking "outside" or "inside" at specific instruments. However, different positions of these AOIs in the cockpit of the aircraft as well as dynamic lighting conditions (e.g., direct sunlight, glare, shadows, and reflections) present challenges for image segmentation [Binaee et al. 2021].

The primary goal of this work is to develop and evaluate a system to capture and analyze a pilot's visual scanning behavior in a real-world *Visual Flight Rules (VFR)* cockpit environment. VFR refers to flying an aircraft primarily using visual references outside the cockpit, typically in clear weather conditions; pilots are expected to "See and Avoid" other aircraft and dedicate most of their attention to visually monitoring the aircraft's surroundings. The flight instruments in the cockpit still play an important role, particularly during other operations such as take-off and landing. To segment the visual field of pilots, we developed a system using Pupil Core eye-tracking glasses. To map pilot's gaze to specific AOIs we evaluated different image segmentation methods including a baseline using OpenCV for coarse "inside vs. outside" dashboard segmentation, as well as more advanced deep learning approaches (Detectron2[2], *SAM 2*[3], *SAM 3*[4]) for instrument-level segmentation.

2 METHODS

The study was conducted in an Aquila AT01 aircraft that features a combination of analog and digital (i.e., glass) cockpits. The Pupil Core glasses (200Hz eye cameras, 60fps scene camera) [Kassner et al. 2014] were used to capture the gaze data of the first author. We evaluated three segmentation architectures to map gaze points to AOIs. Initially, a Raspberry Pi running a custom-compiled version of Pupil Capture for Debian on ARM64 was used, in conjunction with a custom-built iOS app to control the Network API and an HTTP wrapper. However, due to high computational requirements, in data acquisition, we used a laptop which also allowed a quick screen marker calibration instead of the physical markers.

2.1 Segmentation Approaches

To map the pilot's gaze to specific AOIs, the video feed from the scene camera must be analyzed. This process involves segmenting the video frames to distinguish between the pilot looking "outside" (the primary task in VFR flight) and "inside" at specific instruments. The highly dynamic and complex lighting conditions of a cockpit produce various types of artifacts complicate image segmentation. From our experience, it seems that especially shadows thrown inside analog cockpit instruments pose a significant challenge for simpler segmentation methods. To address this, a progressive evaluation was undertaken, moving from a simple baseline model to more advanced architectures.

(1) Dashboard-level (CV2): To distinguish the darker dashboard in the cockpit from the bright outside scene and provide a coarse "inside vs. outside" gaze ratio, we used a baseline model utilizing Gaussian blur and binary thresholding [Bradski 2000].

(2) Instrument-level (Detectron2): While the "inside vs. outside" ratio is useful, additional segmentation is needed to identify which specific instrument a pilot is monitoring. A Mask R-CNN model fine-tuned on a manually annotated dataset of the Aquila AT-01 cockpit [Wu et al. 2019]. This targeted individual instruments with the help of predefined class labels.

(3) Human-in-the-Loop (*SAM 2*): Detectron2 required time intensive manual-annotation and training. Additionally, the model is primarily designed for static images and does not inherently perform object tracking or re-identification across video frames, especially when an instrument is temporarily occluded. To address these limitations, the evaluation progressed using *Segment Anything Model 2 (SAM 2)* [Ravi et al. 2024]. While SAM 2 relies only on visual prompts (manual-click) we also used the *SAM 3* which allows an instructor to define AOIs via visual or text-based prompts on a single frame, leveraging the model's zero-shot generalization for an accurate tracking of instruments across the video. SAM models are not trained on any specific layout and is therefore applicable to a wide variety of aircraft and cockpit configurations. These models are suitable for *human-in-the-loop* analysis: An instructor, without any previous expertise in machine learning, could load a flight video and interactively define AOIs. By clicking on the Airspeed Indicator, Attitude Indicator, and Altimeter, they could generate individual masks for those instruments and immediately proceed with a quantitative analysis of the pilot's scanning patterns.

[2]See https://ai.meta.com/tools/detectron2/
[3]See https://ai.meta.com/research/sam2/
[4]See https://ai.meta.com/research/sam3/

3 EVALUATION

We evaluated our system from two complementary perspectives. First, a *technical evaluation* provides a critical analysis of the system's performance, robustness, and limitations, focusing in particular on the various image segmentation models used for gaze-mapping. Second, a *qualitative evaluation* examines the applicability and perceived usefulness of our system within a real flight training context, drawing on structured feedback from domain experts.

3.1 Technical Evaluation

In ***Dashboard-level Segmentation***, the CV2 achieved real-time processing at 30 FPS on modern hardware. It was robust enough to work across several different analog and digital cockpit types with only minor calibration. However, its reliance on a static color threshold was a primary limitation. CV2 failed in high angle-of-attack (AOA) scenarios, where the bright sky is no longer visible, causing the camera to adjust exposure and when flying over dark terrains such as forests, which were mistakenly classified as "inside". In ***Instrument-level Segmentation*** (see Figure 1), Detectron2 model struggled with difficult lighting conditions and tracking instruments reliably. The trained model is highly specialized for the Aquila AT-01 cockpit and did not perform well in a different aircraft cockpit. This makes the model unsuitable for a general-purpose tool without significant labor-intensive retraining for different aircraft types. On the other hand, SAM 2 and SAM 3 were evaluated as a human-in-the-loop tool that can be used in analog and digital cockpits. SAM 2 still showed some tracking issues in challenging lighting environments, as well as when the instruments go out and in of frame for a certain period. However, SAM 3 was found to be more robust than the other approaches. Another limitation was observed when testing on glass cockpits: the aliasing and shimmering of digital instrument readouts can cause *SAM 3* to produce incomplete AOIs suggesting greater reliability on analog instruments.

3.2 Qualitative Evaluation

In separate sessions, we asked two experts for their subjective feedback to evaluate the practical utility of the system and capture their nuanced opinions. The cohort included active VFR flight instructors, one with extensive prior experience as a military fighter pilot. During these sessions, these experts were presented with the functional prototype.

The most significant finding was the validation of gaze pattern analysis as a valuable tool in aviation. One of the experts noted that during the training pilots' gaze patterns are manually analyzed. This manual and resource-intensive procedure highlights the value placed on scan patterns as an indicator of performance and situational awareness. Hence, our system is recognized as a powerful tool for the automation and quantification of existing analysis. One instructor further stated that the ability to "objectively show a student their scan pattern during a debrief is a powerful tool, far more effective than just telling them they are 'fixating'". The other confirmed that showing a student a replay with their gaze-plot is a "different level of feedback" than simply telling them they are fixating. The consensus was that the tool would be most effective for correcting common but critical VFR errors, such as: a) Fixation on a single instrument during final approach, b) Failure to maintain a visual scan for other traffic (i.e., too much "inside" time), and c) Incomplete or incorrect scan patterns during procedural checks.

In addition, they emphasized that the benefit must be weighed against the overhead of the data capture process. Managing the Pupil Core, calibration, and a laptop in a cramped cockpit is a notable drawback. For the tool to be adopted, data must be presented simply. The "inside vs. outside" gaze ratio was deemed useful due to its simplicity, providing a high-level summary of a student's situational awareness.

4 CONCLUSIONS AND FUTURE WORK

To provide student pilots with objective feedback in post-flight debriefing, we developed a system to capture and analyze their visual scanning behavior in a real-world VFR cockpit environment. In this system, we tested the feasibility of existing scene segmentation models for gaze-mapping. Finally, we conducted a preliminary technical and qualitative evaluation and documented our findings. We are anticipating to obtain constructive feedback and suggestions in the Late Breaking Work session.

REFERENCES

Kamran Binaee, Christian Sinnott, Kaylie Jacleen Capurro, Paul MacNeilage, and Mark D Lescroart. 2021. Pupil Tracking Under Direct Sunlight. In *ACM Symposium on Eye Tracking Research and Applications* (Virtual Event, Germany) *(ETRA '21 Adjunct)*. Association for Computing Machinery, New York, NY, USA, Article 18, 4 pages. https://doi.org/10.1145/3450341.3458490

Gary Bradski. 2000. The opencv library. *Dr. Dobb's Journal: Software Tools for the Professional Programmer* 25, 11 (2000), 120–123.

Andrew T Duchowski. 2018. Gaze-based interaction: A 30 year retrospective. *Computers & Graphics* 73 (2018), 59–69.

Flavio Ferrari, Kevin PC Spillmann, Chiara P Knecht, Kenan Bektaş, and Celine M Muehlethaler. 2018. Improved pilot training using head and eye tracking system. In *Eye Tracking for Spatial Research, Proceedings of the 3rd International Workshop (ET4S)*. ETH Zurich, 58–63.

PM Fitts, RE Jones, and JL Milton. 1950. Eye movements of aircraft pilots during instrument-landing approaches. Aeronaut. Eng. Rev. (1950).

Moritz Kassner, William Patera, and Andreas Bulling. 2014. Pupil: an open source platform for pervasive eye tracking and mobile gaze-based interaction. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication* (Seattle, Washington) *(UbiComp '14 Adjunct)*. Association for Computing Machinery, New York, NY, USA, 1151–1160. https://doi.org/10.1145/2638728.2641695

Diederick C. Niehorster, Michael Hildebrandt, Anthony Smoker, Halszka Jarodzka, and Nicklas Dahlström. 2020. *Towards eye tracking as a support tool for pilot training and assessment.* 17–28 pages. https://doi.org/10.3929/ethz-b-000407625

Nikhila Ravi, Valentin Gabeur, Yuan-Ting Hu, Ronghang Hu, Chaitanya Ryali, Tengyu Ma, Haitham Khedr, Roman Rädle, Chloe Rolland, Laura Gustafson, Eric Mintun, Junting Pan, Kalyan Vasudev Alwala, Nicolas Carion, Chao-Yuan Wu, Ross Girshick, Piotr Dollár, and Christoph Feichtenhofer. 2024. SAM 2: Segment Anything in Images and Videos. *arXiv preprint arXiv:2408.00714* (2024). https://arxiv.org/abs/2408.00714

Yuxin Wu, Alexander Kirillov, Francisco Massa, Wan-Yen Lo, and Ross Girshick. 2019. Detectron2. https://github.com/facebookresearch/detectron2.